

FIGURE 1

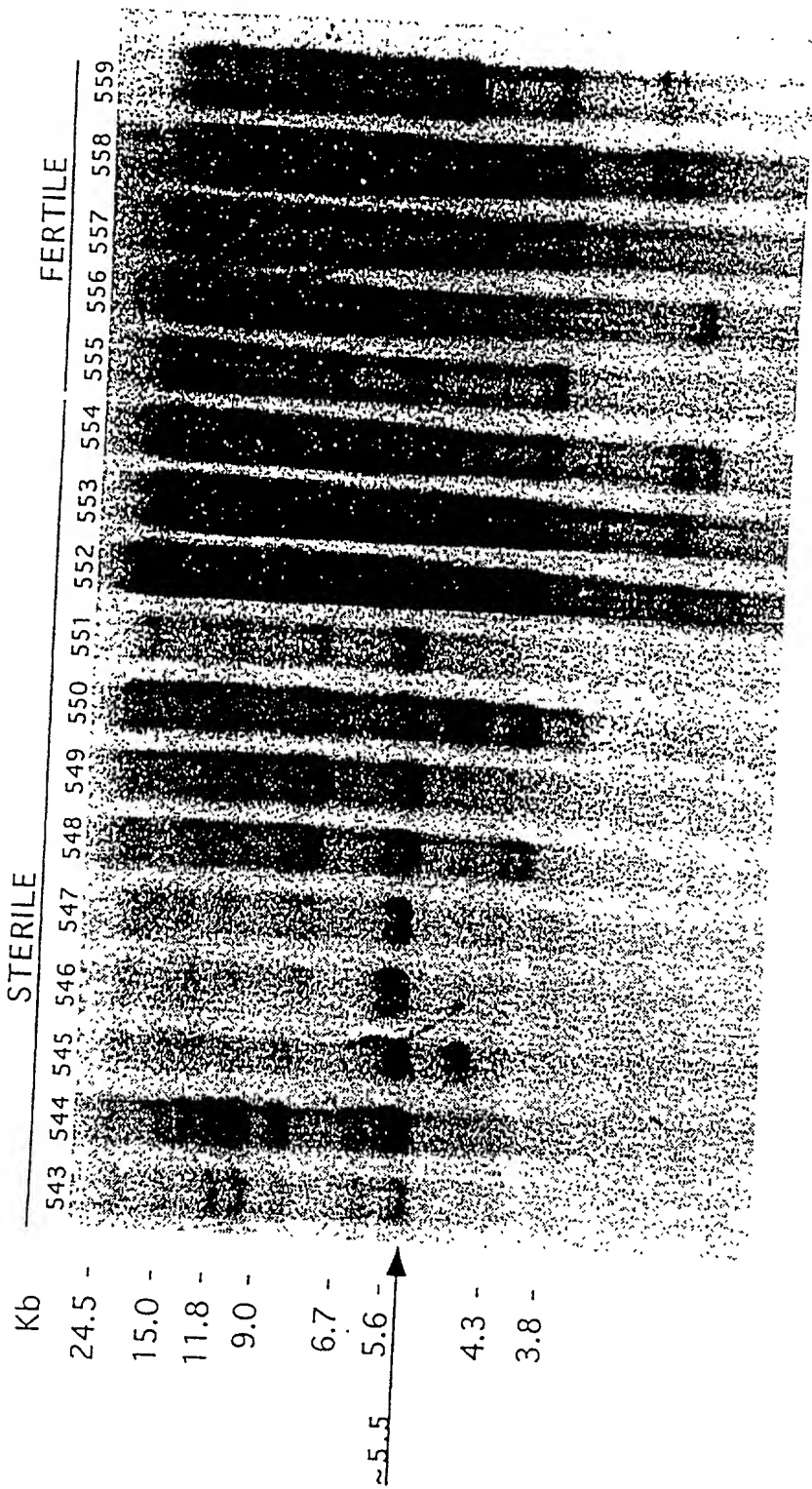


FIGURE 2

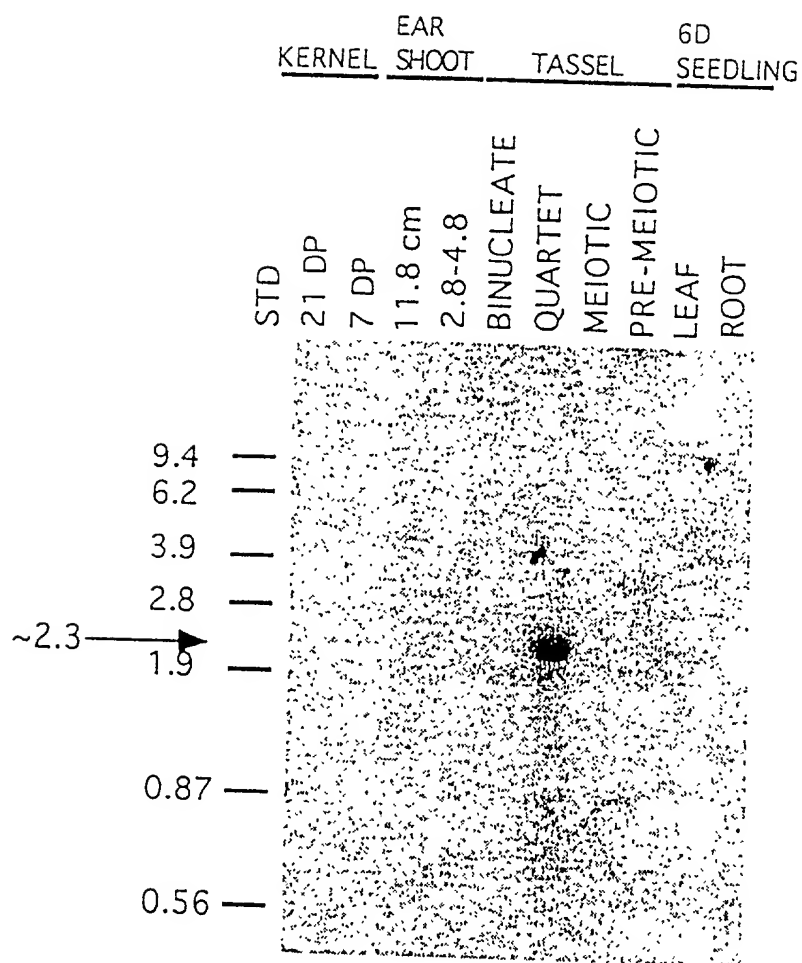


FIGURE 3

Figure 4

EcoRI
|
GAATTCGGCACGAGGGAAGCTCACCTCACGCCGGCGACGCCATCGCCATTCTTCCCCTA
1 -----+-----+-----+-----+-----+-----+ 60
CTTAAGCCGTGCTCCCTTCGAGTGGAGTGCGCCGCTGCGGTAGCGGTAAGAAGGGTGAT

a E F G T R E A H L T P A T P S P F F P L -

GCAGGGCCTCACAAGTACATCGCGCTCCTTCTGGTTGTCCTCTCATGGATCCTGGTCCAG
61 -----+-----+-----+-----+-----+-----+ 120
CGTCCCGGAGTGTTTCATGTAGCGCGAGGAAGACCAACAGGAGAGTACCTAGGACCAGGTC

a A G P H K Y I A L L L V V L S W I L V Q -

AGGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGCGCAACGGTG
121 -----+-----+-----+-----+-----+-----+ 180
TCCACCTCGGACTCCTTCGTCTTTCCGGGCTCTAGTACCGGTACAGTACCGCGGTTGCCAC

a R W S L R K Q K G P R S W P V I G A T V -

GAGCAGCTGAGGAACCTACCAACGGATGCACGACTGGCTTGTCGGGTACCTGTCACGGCAC
181 -----+-----+-----+-----+-----+-----+ 240
CTCGTCGACTCCTTGATGGTGGCCTACGTGCTGACCGAACAGCCCATGGACAGTGCCGTG

a E Q L R N Y H R M H D W L V G Y L S R H -

AGGACAGTGACCGTCGACATGCCGTTCACTTCCTACACCTACATCGCTGACCCGGTGAAT
241 -----+-----+-----+-----+-----+-----+ 300
TCCTGTCACTGGCAGCTGTACGGCAAGTGAAGGATGTGGATGTAGCGACTGGGCCACTTA

a R T V T V D M P F T S Y T Y I A D P V N -

GTCGAGCATGTCCTCAAGACTAACTTCACCAATTACCCCAAGGGAATCGTGACAGATCC
301 -----+-----+-----+-----+-----+-----+ 360
CAGCTCGTACAGGAGTTCTGATTGAAGTGGTTAATGGGGTTCCTTAGCACATGTCTAGG

a V E H V L K T N F T N Y P K G I V Y R S -

TACATGGACGTGCTCCTCGGTGACGGCATCTTCAACGCCGACGGCGAGCTGTGGAGGAAG
361 -----+-----+-----+-----+-----+-----+ 420
ATGTACCTGCACGAGGAGCCACTGCCGTAGAAGTTGCGGCTGCCGCTCGACACCTCCTTC

a Y M D V L L G D G I F N A D G E L W R K -

CAGAGGAAGACGGCGAGTTTCGAGTTCGCCTCCAAGAACCTGAGGGATTTTCAGCGCCATT
421 -----+-----+-----+-----+-----+-----+ 480
GTCTCCTTCTGCCGCTCAAAGCTCAAGCGGAGGTTCTTGACTCCCTAAAGTCGCGGTAA

a Q R K T A S F E F A S K N L R D F S A I -

100157-1244

Figure 4B

GTGTTTCAGAGAGTACTCCCTGAAGCTGTCTGGGTATACTGAGCCAGGCATCCAAGGCAGGC
 481 -----+-----+-----+-----+-----+-----+-----+ 540
 CACAAGTCTCTCATGAGGGACTTCGACAGCCCATATGACTCGGTCCGTAGGTTCCGTCCG
 a V F R E Y S L K L S G I L S Q A S K A G -
 AAAGTTGTGGACATGCAGGAACCTTTACATGAGGATGACGCTGGACTCCATCTGCAAGGTT
 541 -----+-----+-----+-----+-----+-----+-----+ 600
 TTTCAACACCTGTACGTCCTTGAAATGTACTCTACTGCGACCTGAGGTAGACGTTCCAA
 a K V V D M Q E L Y M R M T L D S I C K V -
 GGGTTCGGGGTCGAGATCGGCACGCTGTCTGCCAGATCTCCCCGAGAACAGCTTCGCGCAG
 601 -----+-----+-----+-----+-----+-----+-----+ 660
 CCCAAGCCCCAGCTCTAGCCGTGCGACAGCGGTCTAGAGGGGCTCTTGTCGAAGCGCGTC
 a G F G V E I G T L S P D L P E N S F A Q -
 GCGTTCGATGCCCAACATCATCATCACGCTGCGGTTTCATCGACCCGCTGTGGCGCATC
 661 -----+-----+-----+-----+-----+-----+-----+ 720
 CGCAAGCTACGGCGGTTGTAGTAGTAGTGCAGCGCCAAGTAGCTGGGCGACACCGCGTAG
 a A F D A A N I I I T L R F I D P L W R I -
 AAGAGGTTCTTCCACGTCGGGTCTAGAGGCCCTCCTAGCGCAGAGCATCAAGCTCGTGGAC
 721 -----+-----+-----+-----+-----+-----+-----+ 780
 TTCTCCAAGAAGGTGCAGCCAGTCTCCGGGAGGATCGCGTCTCGTAGTTCGAGCACCTG
 a K R F F H V G S E A L L A Q S I K L V D -
 GAGTTCACCTACAGCGTGATCCGCCGAGGAAGGCCGAGATCGTTCGAGGTCCGGGCCAGC
 781 -----+-----+-----+-----+-----+-----+-----+ 840
 CTCAAGTGGATGTCGACTAGGCGGCCTCCTTCCGGCTCTAGCAGCTCCAGGCCCGGTGCG
 a E F T Y S V I R R R K A E I V E V R A S -
 GGCAAACAGGAGAAGATGAAGCACGACATCCTGTCTACGGTTCATCGAGCTGGGCGAGGCC
 841 -----+-----+-----+-----+-----+-----+-----+ 900
 CCGTTTGTCTCTTCTACTTCGTGCTGTAGGACAGTGCCAAGTAGCTCGACCCGCTCCGG
 a G K Q E K M K H D I L S R F I E L G E A -
 GGCGACGACGGCGGCGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTGCTCAACTTC
 901 -----+-----+-----+-----+-----+-----+-----+ 960
 CCGCTGCTGCCGCCGCCGAAGCCCCTGCTATTCTCGGAGGCCCTGCACCACGAGTTGAAG
 a G D D G G G F G D D K S L R D V V L N F -
 GTGATCGCCGGGCGGGACACGACGGCGACGCTGTCTGTTGTTACGCACATGGCCATG
 961 -----+-----+-----+-----+-----+-----+-----+ 1020
 CACTAGCGGCCCCGCCCTGTGCTGCGCTGCTGCGACAGCACCAAGTGCGTGTACCGGTAC
 a V I A G R D T T A T T L S W F T H M A M -

bioRxiv preprint doi: <https://doi.org/10.1101/100000>; this version posted October 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4C

```

TCCCACCCGGACGTGGCCGAGAAGCTGCGCCGCGAGCTGTGCGCGTTCGAGGCGGAGCGC
1021 -----+-----+-----+-----+-----+ 1080
AGGGTGGGCCTGCACCGGCTCTTCGACGCGGCGCTCGACACGCGCAAGCTCCGCCTCGCG

a      S H P D V A E K L R R E L C A F E A E R -

GCGCGCGAGGAGGGCGTCACGCTCGTGCTCTGCGGCGGCGCTGACGCCGACGACAAGGCG
1081 -----+-----+-----+-----+-----+ 1140
CGCGCGCTCCTCCCGCAGTGCAGACACGAGACGCCGCCGCGACTGCGGCTGTGTTCCGC

a      A R E E G V T L V L C G G A D A D D K A -

TTCGCCGCCCGCGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTG
1141 -----+-----+-----+-----+-----+ 1200
AAGCGGCGGGCGCACCGCGTCAAGCGCCCGGAGGAGTGGATGCTGTGCGAGCCGTTTCGAC

a      F A A R V A Q F A G L L T Y D S L G K L -

GTCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTCCCTCAGGAC
1201 -----+-----+-----+-----+-----+ 1260
CAGATGGAGGTGCGGACGCACTGGCTCTGCGAGGCGGACATGGGGCGGCAGGAGTCTCTG

a      V Y L H A C V T E T L R L Y P A V P Q D -

CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAGGGCCGGCGGG
1261 -----+-----+-----+-----+-----+ 1320
GGGTTCCTCCCTAGGACCTCCTGCTGCACGACGGCCTGCCCTGCTTCCACTCCCGGCCGCC

a      P K G I L E D D V L P D G T K V R A G G -

ATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGTACAACCTGGGGCCCCGACGCG
1321 -----+-----+-----+-----+-----+ 1380
TACCACTGCATGCACGGGATGAGCTACCCCGCCTACCTCATGTTGACCCCGGGGCTGCGC

a      M V T Y V P Y S M G R M E Y N W G P D A -

GCGAGCTTCCGGCCGGAGCGGTGGATCAACGAGGATGGCGCGTTCGCAACGCGTCGCCC
1381 -----+-----+-----+-----+-----+ 1440
CGCTCGAAGGCCGGCCTCGCCACCTAGTTGCTCCTACCGCGCAAGGCGTTGCGCAGCGGC

a      A S F R P E R W I N E D G A F R N A S P -

TTCAAGTTCACGGCGTTCCAGGCGGGGCCGAGGATCTGCCTGGGCAAGGACTCGGCGTAC
1441 -----+-----+-----+-----+-----+ 1500
AAGTTCAAGTGCCGCAAGGTCCGCCCCGGCTCCTAGACGGACCCGTTCTGAGCCGCATG

a      F K F T A F Q A G P R I C L G K D S A Y -

CTGCAGATGAAGATGGCGCTGGCCATCCTCTTCCGCTTCTACAGCTTCCGGCTGCTGGAG
1501 -----+-----+-----+-----+-----+ 1560
GACGTCTACTTCTACCGCGACCGGTAGGAGAAGGCGAAGATGTGCAAGGCCGACGACCTC

a      L Q M K M A L A I L F R F Y S F R L L E -

```

10001657-15140

Figure 4D

GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGGCCTCAAGGTC
1561 -----+-----+-----+-----+-----+ 1620
CCCGTGGGCCACGTCATGGCGTACTACTGGTAGGAGAGGTACCGCGTGCCGGAGTTCCAG

a G H P V Q Y R M M T I L S M A H G L K V -

CGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTGGATATGGATATCGTCCCGCTTAAT
1621 -----+-----+-----+-----+-----+ 1680
GCGCAGAGATCCCGGCAGACTACAGTACCGCTAAACCTATACCTATAGCAGGGCGAATTA

a R V S R A V * C H G D L D M D I V P L N -

CCACGACAAATAACGCTCGTGTTACAAATTTGCATGCATGCATGTAAGGGAAAGCGATGG
1681 -----+-----+-----+-----+-----+ 1740
GGTGCTGTTTATTGCGAGCACAATGTTTAAACGTACGTACGTACATTCCCTTTCGCTACC

a P R Q I T L V L Q I C M H A C K G K R W -

GTTTCATTGGTGGCTTGGCTTAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACA
1741 -----+-----+-----+-----+-----+ 1800
CAAAGTAACCACCGAACCGAATTCGGAATTTTGGAGGCAGCCAGAACGCTTGGTGGTGT

a V S L V A W L K P *

TCACTAGTGTTTTGTACTCTACTCCTCAGTGGAAGTGTAGTGACAGCATACAAGTTCATC
1801 -----+-----+-----+-----+-----+ 1860
AGTGATCACAAAACATGAGATGAGGAGTCACCTTCACATCACTGTCGTATGTTCAAGTAG

XhoI
|
ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAAAACTCGAG
1861 -----+-----+-----+-----+----- 1906
TATATATAATAGGAGAAAGAATTTTTTTTTTTTTTTTTTTTGGAGCTC

10031657.24404

Figure 5

1 GAATTCCAAG CGAGGCCCTT GTAGCAGAGA GTGTTGCTGA TGCAGTCGGC
51 GGAAATGAGT GCGTGCTGAG AGCAACGCTG AGGGGTTCCTA GGGATGGCAA
101 TGGCTATGGC AATCGGCTAG AGGTGGAGGA CAAGGTGGTG AGGATTGGGA
151 GGGCAACCTA TGGCAAGTTG GTGAAGAGGC ACGCAATGAG AGATCTATTC
201 AGACTTACAC TGGATGCCGC CAACAAATTC AACCTTTAGA TTTTGATACT
251 GTCACCTCCTA CTTTATTCCT TGGTTGGGCA ACTTCCAATA GGCTCATGTT
301 AATCAATGAT TAGTGATTAT TCAGCAAATA TTCTTGTTTG TTTGACATTT
351 ATAATATGTG GGGTGAGACG GATTAAATAT CATCCATGAG AGCTTTATCT
401 TCATGCTCTC TTGATTTTGG TTTCAGATCA TTCTTTCAGT GTTCACAAGA
451 ATTTTCTCAG TTTGGTCCAT GTAATTTTGG AAGTGAGGTT CCTTAAATTT
501 CATTATGCTT CCTTCTTTT CTAGACTAGC AACTGCATGA CTTTTCACCT
551 TGGGTTTACA AATTGACTCA CAAGAAAACA AATTCACCTT TGGGTTTACA
601 AATTCCTCTT CAGGATGTAC TTTTCACTTG AACTGTCATG TATAGGAACA
651 AGGAATGGCT CAGTTTTTAA GGAACAATGT ACAGATTTCA TTTCAGAACT
701 CTTTCTGGTT GGTGAGTTT CAGACTTTTT GTACCAAGCT GATGGATCAC
751 AATACTTGTT TCCAAAGTCT GATAACAGAA ACTGGCAACT CCTAATTGAT
801 AATAAAAAGA ATAAAATACA GTATCAGATA TCTCATTTTC TTGGTTGGCA
851 GATCACAAAA AGGAACACAA AGGCTAAGCC TCCTACTTGT TCGGGAGTTA
901 GGTCAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTTGG GGTCACACCA
951 AGATTGTCTC TCTCGAGGTT GGGGGGTCCC TAAGGTTGGT AGTAGCAATA
1001 CCCAATATAT CACCTAACAA ACCCAATCCA TGCTACATAC ATACATAGCA
1051 TCCATCACTT GTAGACTGGA CCCTTCATCA AGAGCACCAT GGAGGAAGCT
1101 CACATCACGC CGGCGACGCC ATCGCCATTC TTCCCCTAG CAGGGCCTCA
1151 CAAGTACATC GCGCTCCTCC TGGTTGTCCT CTCATGGATC CTGGTCCAGA
1201 GGTGGAGCCT GAGGAAGCAG AAAGGCCCGA GATCATGGCC AGTCATCGGT
1251 GCAACGGTGG AGCAGCTGAG GAACTACCAC CGGATGCACG ACTGGCTTGT
1301 CGGGTACCTG TCACGGCACA GGACAGTGAC CGTCGACATG CCGTTCACTT
1351 CCTACACCTA CATCGCTGAC CCGGTGAATG TCGAGCATGT CCTCAAGACT

Figure 5B

1401 AACTTCACCA ATTACCCCAA GGTAAATGAC CTGAACTCAC TGATGTTTCAG
1451 TCTTCGGAAA TCAGAGCTGA AAGCTGAATC GAATGTGCCT GAACACCGTG
1501 TAGGGAATCG TGTACAGATC CTACATGGAC GTGCTCCTCG GTGACGGCAT
1551 CTTCAACGCC GACGGCGAGC TGTGGAGGAA GCAGAGGAAG ACGGCGAGTT
1601 TCGAGTTCGC CTCCAAGAAC CTGAGGGATT TCAGCGCCAT TGTGTTTCAGA
1651 GAGTACTCCC TGAAGCTGTC GGGTATACTG AGCCAGGCAT CCAAGGCAGG
1701 CAAAGTTGTG GACATGCAGG TGAGATCACT GCTCCCTTGC CATTGCCAAC
1751 ATGAGCATTT CAACCTGAGA CACGAGAGCT ACCTTGCCGA TTCAGGAACT
1801 TTACATGAGG ATGACGCTGG ACTCCATCTG CAAGGTTGGG TTCGGGGTCG
1851 AGATCGGCAC GCTGTCGCCG GATCTCCCCG AGAACAGCTT CGCGCAGGCG
1901 TTCGATGCCG CCAACATCAT CGTCACGCTG CGGTTTCATCG ACCCGCTGTG
1951 GCGCATCAAG AGGTTCTTCC ACGTCGGGTC AGAGGCCCTC CTAGCGCAGA
2001 GCATCAAGCT CGTGGACGAG TTCACCTACA GCGTGATCCG CCGGAGGAAG
2051 GCCGAGATCG TCGAGGCCCG GGCCAGCGGC AAACAGGAGA AGGTACGTGC
2101 ACATGACTGT TTCGATTCTT CAGTTCATCG TCTTGGCCGG GATGGACCTG
2151 ATCCTGATTG ATTATATATC CGTGTGACTT GTGAGGACAA ATTAATATGG
2201 GCAGATGAAG CACGACATCC TGTCACGGTT CATCGAGCTA GGCGAGGCCG
2251 GCGACGACGG CGGCGGCTTC GGGGACGACA AGAGCCTCCG GGACGTGGTG
2301 CTCAACTTCG TGATCGCCGG GCGGGACACG ACGGCGACGA CGCTGTCTGTG
2351 GTTCACGCAC ATGGCCATGT CCCACCCGGA CGTGGCCGAG AAGCTGCGCC
2401 GCGAGCTGTG CGCGTTCGAG GCGGAGCGCG CGCGCGAGGA GGGCGTCCGG
2451 CTCGTGCCCT GCGGCGGCGC TGACGCCGAC GACAAGGCGT TCGCCGCCCCG
2501 CGTGGCGCAG TTCGCGGGCC TCCTCACCTA CGACAGCCTC GGCAAGCTGG
2551 TCTACCTCCA CGCCTGCGTC ACCGAGACGC TCCGCCTGTA CCCC GCCGTC
2601 CCTCAGGTGA GCGCGCCCGA CACGCGACCT CCGGTCCAGA GCACAGCATG
2651 CAGTGAGTGG ACCTGAATGC AATGCACATG CACTTGCGCG CGCGCAGGAC
2701 CCCAAGGGGA TCCTGGAGGA CGACGTGCTG CCGGACGGGA CGAAGGTGAG
2751 GGCCGGCGGG ATGGTGACGT ACGTGCCCTA CTCGATGGGG CGGATGGAGT

Figure 5C

2801 ACAACTGGGG CCCCACGCG GCGAGCTTCC GGCCGGAGCG GTGGATCAAC
2851 GAGGATGGCG CGTTCCGCAA CGCGTCGCCG TTCAAGTTCA CGGCGTTCCA
2901 GGCGGGGCCG AGGATCTGCC TGGGCAAGGA CTCGGCGTAC CTGCAGATGA
2951 AGATGGCGCT GGCCATCCTC TTGCGCTTCT ACAGCTTCCG GCTGCTGGAG
3001 GGGCACCCGG TGCAGTACCG CATGATGACC ATCCTCTCCA TGGCGCACGG
3051 CCTCAAGGTC CGCGTCTCTA GGGCCGTCTG ATGTCATGGC GATTTGGGAT
3101 ATCATCCCGC TTAATCCTTA AAAATTTGCA TGCATGCATG TAAGGGAAAG
3151 CGATGGGTTT CATTGGTGGC TTGGCTTAAG CCTTAAAAAC TCCGTCGGGT
3201 CTTGCGAACC ACCACATCAC TAGTGTTTTG TACTCTACTC CTCAGTGGAA
3251 GTGTAGTGAC AGCATACAAG TTCATCATAT ATATTATCCT CTTTCTTCG
3301 CGGATGCTTC CCGGGACCTT TTGGAGACCA TTACTGACAG GCGTGTGAAA
3351 AAAAGGCTTC TTCTGCGGCG AAGTTTTGGG TTCAGAGTCT TGGCGTCTTT
3401 GCAGCAGAAA AAAGGTTTGG AAGGATCTGA ACCCTGAACC GAAAATGGCT
3451 TCGGAAATAT GCTCGCATCG GGGCGGGGCC GTCACTCGGG ATGACGACAA
3501 GCCCACAAGC AGTGAGAGCG AAGCGATCTT TGGAGTTTGG AGACACTCTC
3551 GGACCCCTCG GCGCTCCGCG AGCTCATCTT CGCCTCCTCT GTCGTGTCCG
3601 TGGCGGCACC GCGCCCGCCC GCCTCGTGTT CGACCAAATC CCGCGCCCCG
3651 ACCGTTTCGT GTACAACACC CTCATCCGCG GCGCCGCGCG CAGTGACACG
3701 CCCCAGGACG CCGTATACAT CTATAAATCA TGGTATTGTA CTTTATTTTC
3751 AAACGGCCTT AACACAACCA TATTTTATG GTAAACACGT TCAAAATTGA
3801 CACAAATTTA AAACAGGCAC AAACCGTAGC TAAACATAAG AGAATGAGAG
3851 ACAACCCAAA GGTTAGAGAT GAAATAAGCT GAGTAAACGA CGAATTC

Figure 6

1051 TCCATCACTTGTAGACTGGACCCCTTCATCAAGAGCACCATGGAGGAAGCT 1100
1GAATTCGGCACGAGGGAAGCT 21
1101 CACATCACGCCGGCGACGCCATCGCCATTCTTCCCACTAGCAGGGCCTCA 1150
22 CACCTCACGCCGGCGACGCCATCGCCATTCTTCCCACTAGCAGGGCCTCA 71
1151 CAAGTACATCGCGCTCCTCTCGTTGTCCTCTCATGGATCCTGGTCCAGA 1200
72 CAAGTACATCGCGCTCCTCTCGTTGTCCTCTCATGGATCCTGGTCCAGA 121
1201 GGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGT 1250
122 GGTGGAGCCTGAGGAAGCAGAAAGGCCCGAGATCATGGCCAGTCATCGGC 171
1251 GCAACGGTGGAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGT 1300
172 GCAACGGTGGAGCAGCTGAGGAACTACCACCGGATGCACGACTGGCTTGT 221
1301 CGGGTACCTGTACGGCACAGGACAGTGACCGTCGACATGCCGTTCACTT 1350
222 CGGGTACCTGTACGGCACAGGACAGTGACCGTCGACATGCCGTTCACTT 271
1351 CCTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAGACT 1400
272 CCTACACCTACATCGCTGACCCGGTGAATGTCGAGCATGTCCTCAAGACT 321
1401 AACTTCACCAATTACCCCAAGGTAAATGACCTGAACTCACTGATGTTTCA 1450
322 AACTTCACCAATTACCCCA..... 340

1501 TAGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 1550
341 .AGGGAATCGTGTACAGATCCTACATGGACGTGCTCCTCGGTGACGGCAT 389
1551 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 1600
390 CTTCAACGCCGACGGCGAGCTGTGGAGGAAGCAGAGGAAGACGGCGAGTT 439
1601 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTAGCGCCATTGTGTTCAGA 1650
440 TCGAGTTCGCCTCCAAGAACCTGAGGGATTTAGCGCCATTGTGTTCAGA 489
1651 GAGTACTCCCTGAAGCTGTGCGGTATACTGAGCCAGGCATCCAAGGCAGG 1700
490 GAGTACTCCCTGAAGCTGTGCGGTATACTGAGCCAGGCATCCAAGGCAGG 539
1701 CAAAGTTGTGGACATGCAGGTGAGATCACTGCTCCCTTGCCATTGCCAAC 1750
540 CAAAGTTGTGGACATG..... 555

Figure 6B

```

1751 ATGAGCATTTCAACCTGAGACACGAGAGCTACCTTGCCGATTCAGGAACT 1800
      |||||
556 .....CAGGAACT 563

1801 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTCTGGGGTCG 1850
      |||||
564 TTACATGAGGATGACGCTGGACTCCATCTGCAAGGTTGGGTTCTGGGGTCG 613

1851 AGATCGGCACGCTGTCTGCCGGATCTCCCCGAGAACAGCTTCGCGCAGGCG 1900
      |||||
614 AGATCGGCACGCTGTCTGCCAGATCTCCCCGAGAACAGCTTCGCGCAGGCG 663

1901 TTCGATGCCGCCAACATCATCTGTCACGCTGCGGTTTCATCGACCCGCTGTG 1950
      |||||
664 TTCGATGCCGCCAACATCATCTGTCACGCTGCGGTTTCATCGACCCGCTGTG 713

1951 GCGCATCAAGAGGTTCTTCCACGTCGGGTCTAGAGGCCCTCCTAGCGCAGA 2000
      |||||
714 GCGCATCAAGAGGTTCTTCCACGTCGGGTCTAGAGGCCCTCCTAGCGCAGA 763

2001 GCATCAAGCTCGTGGACGAGTTACCTACAGCGTGATCCGCCGAGGAAG 2050
      |||||
764 GCATCAAGCTCGTGGACGAGTTACCTACAGCGTGATCCGCCGAGGAAG 813

2051 GCCGAGATCGTCTGAGGCGCGGGCCAGCGGCAAACAGGAGAAGGTACGTGC 2100
      |||||
814 GCCGAGATCGTCTGAGGTCCGGGCGAGCGGCAAACAGGAGA..... 853

      .
      .

2201 GCAGATGAAGCACGACATCTGTACGCGTTTCATCGAGCTAGGCGAGGCCG 2250
      |||||
854 ..AGATGAAGCACGACATCTGTACGCGTTTCATCGAGCTAGGCGAGGCCG 901

2251 GCGACGACGGCGGGCGGCTTCGGGGACGACAAGAGCCTCCGGGACGTGGTG 2300
      |||||
902 GCGACGACGGCGGGCGGCTTCGGGGACGATAAGAGCCTCCGGGACGTGGTG 951

2301 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCTGTG 2350
      |||||
952 CTCAACTTCGTGATCGCCGGGCGGGACACGACGGCGACGACGCTGTCTGTG 1001

2351 GTTCACGCACATGGCCATGTCCCACCCGACGTGGCCGAGAAGCTGCGCC 2400
      |||||
1002 GTTCACGCACATGGCCATGTCCCACCCGACGTGGCCGAGAAGCTGCGCC 1051

2401 GCGAGCTGTGCGCGTTTCGAGGCGGAGCGCGCGCGAGGAGGGCGTCGCG 2450
      |||||
1052 GCGAGCTGTGCGCGTTTCGAGGCGGAGCGCGCGCGAGGAGGGCGTCACG 1101

2451 CTCGTGCCCTGCGGCGGGCGCTGACGCCGACGACAAGGCGTTCCGCCCCCG 2500
      |||||
1102 CTCGTGCTCTGCGGCGGGCGCTGACGCCGACGACAAGGCGTTCCGCCCCCG 1151

```

Figure 6C

2501 CGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTGG 2550
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1152 CGTGGCGCAGTTCGCGGGCCTCCTCACCTACGACAGCCTCGGCAAGCTGG 1201
 2551 TCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTC 2600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1202 TCTACCTCCACGCCTGCGTCACCGAGACGCTCCGCCTGTACCCCGCCGTC 1251
 2601 CCTCAGGTGAGCGCGCCCCACACGCGACCTCCGGTCCAGAGCACAGCATG 2650
 ||||
 1252 CCT..... 1254
 2651 CAGTGAGTGGACCTGAATGCAATGCACATGCACTTGCGCGCGCGCAGGAC 2700
 ||||||
 1255CAGGAC 1260
 2701 CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAG 2750
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1261 CCCAAGGGGATCCTGGAGGACGACGTGCTGCCGGACGGGACGAAGGTGAG 1310
 2751 GGCCGGCGGGATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGT 2800
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1311 GGCCGGCGGGATGGTGACGTACGTGCCCTACTCGATGGGGCGGATGGAGT 1360
 2801 ACAACTGGGGCCCCGACGCGGCGAGCTTCCGGCCGGAGCGGTGGATCAAC 2850
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1361 ACAACTGGGGCCCCGACGCGGCGAGCTTCCGGCCGGAGCGGTGGATCAAC 1410
 2851 GAGGATGGCGCGTTCCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 2900
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1411 GAGGATGGCGCGTTCCGCAACGCGTCGCCGTTCAAGTTCACGGCGTTCCA 1460
 2901 GGCGGGGCGGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 2950
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1461 GGCGGGGCGGAGGATCTGCCTGGGCAAGGACTCGGCGTACCTGCAGATGA 1510
 2951 AGATGGCGCTGGCCATCCTCTTGCCTTCTACAGCTTCCGGCTGCTGGAG 3000
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1511 AGATGGCGCTGGCCATCCTCTTGCCTTCTACAGCTTCCGGCTGCTGGAG 1560
 3001 GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 3050
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1561 GGGCACCCGGTGCAGTACCGCATGATGACCATCCTCTCCATGGCGCACGG 1610
 3051 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTTG.... 3096
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1611 CCTCAAGGTCCGCGTCTCTAGGGCCGTCTGATGTCATGGCGATTTGGATA 1660
 3097 .GGATATCATCCCGCTTAATCC.....TTAAAAATT 3126
 |||||||||
 1661 TGGATATCGTCCCGCTTAATCCACGACAAATAACGCTCGTGTACAAATT 1710
 3127 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 3176
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1711 TGCATGCATGCATGTAAGGGAAAGCGATGGGTTTCATTGGTGGCTTGGCT 1760

Figure 6D

3177 TAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACATCACTAGTGT 3226
|||||
1761 TAAGCCTTAAAACTCCGTCGGGTCTTGCGAACCACCACATCACTAGTGT 1810
3227 TTTGTACTCTACTCCTCAGTGGAGTGTAGTGACAGCATACAAGTTCATC 3276
|||||
1811 TTTGTACTCTACTCCTCAGTGGAGTGTAGTGACAGCATACAAGTTCATC 1860
3277 ATATATATTATCCTCTTTCTTCGCCGGATGCTTCCCGGGACCTTTTGGAG 3326
|||||
1861 ATATATATTATCCTCTTTCTTAAAAAAAAAAAAAAAAAACTCGAG.... 1906
.
.
.

1004657.124404

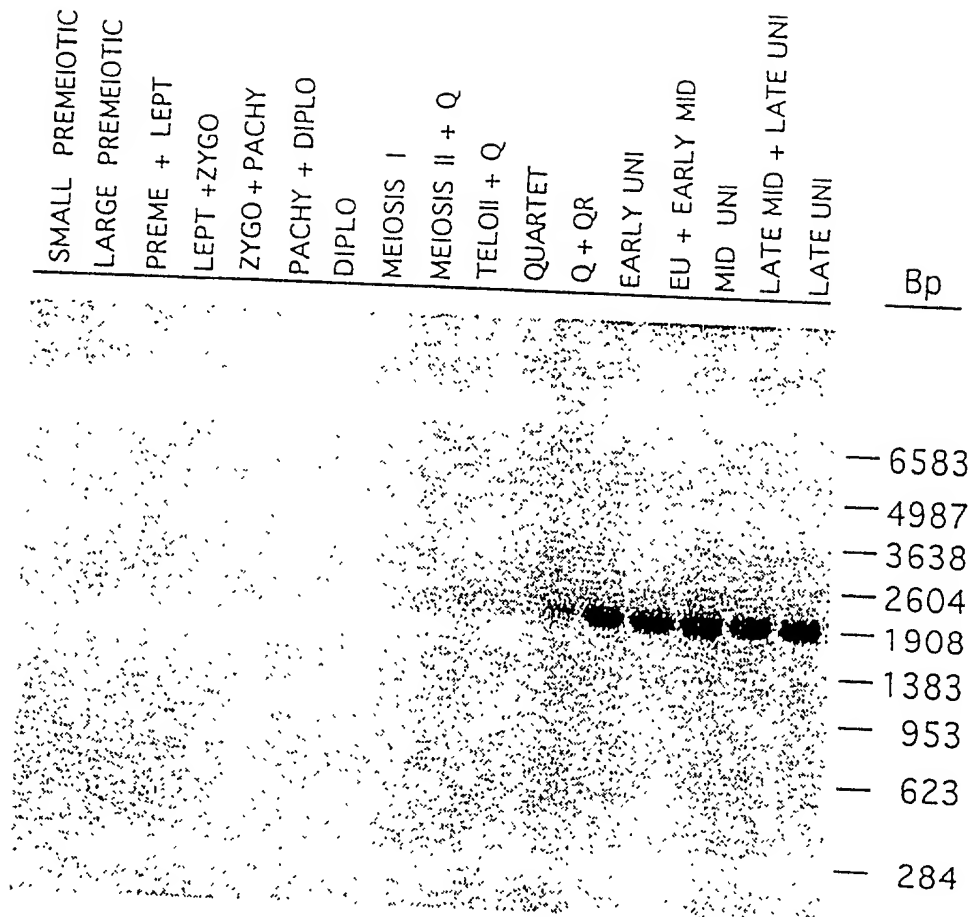


Figure 7

Figure 8

1 GAATTCCAAG CGAGGCCCTT GTAGCAGAGA GTGTTGCTGA TGCAGTCGGC
51 GGAAATGAGT GCGTGCTGAG AGCAACGCTG AGGGGTTCCA GGGATGGCAA
101 TGGCTATGGC AATCGGCTAG AGGTGGAGGA CAAGGTGGTG AGGATTGGGA
151 GGGCAACCTA TGGCAAGTTG GTGAAGAGGC ACGCAATGAG AGATCTATTC
201 AGACTTACAC TGGATGCCGC CAACAAATTC AACCTTTAGA TTTTGATACT
251 GTCACTCCTA CTTTATTCCT TGGTTGGGCA ACTTCCAATA GGCTCATGTT
301 AATCAATGAT TAGTGATTAT TCAGCAAATA TTCTTGTTTG TTTGACATTT
351 ATAATATGTG GGGTGAGACG GATTAAATAT CATCCATGAG AGCTTTATCT
401 TCATGCTCTC TTGATTTTGG TTTCAGATCA TTCTTTCAGT GTTCACAAGA
451 ATTTTCTCAG TTTGGTCCAT GTAATTTTTG AAGTGAGGTT CCTTAAATTT
501 CATTATGCTT CCTTTCTTTT CTAGACTAGC AACTGCATGA CTTTTCACCT
551 TGGGTTTACA AATTGACTCA CAAGAAAACA AATTCACCTT TGGGTTTACA
601 AATTCCTCTT CAGGATGTAC TTTTCACTTG AACTGTCATG TATAGGAACA
651 AGGAATGGCT CAGTTTTTAA GGAACAATGT ACAGATTTCA TTTCAGAACT
701 CTTTCTGGTT GGTTGAGTTT CAGACTTTTT GTACCAAGCT GATGGATCAC
751 AATACTTGTT TCCAAAGTCT GATAACAGAA ACTGGCAACT CCTAATTGAT
801 AATAAAAAGA ATAAAATACA GTATCAGATA TCTCATTTTC TTGGTTGGCA
851 GATCACAAAA AGGAACACAA AGGCTAAGCC TCCTACTTGT TCGGGAGTTA
901 GGTCAGGGAC ACCATATGAA TGAAAGAAAT CTTAATTGG GTTCACACCA
951 AGATTGTCTC TCTCGAGGTT GGGGGGTCCC TAAGGTGGT AGTAGCAATA
1001 CCCAATATAT CACCTAACAA ACCCAATCCA TGCTACATAC ATACATAGCA
1051 TCCATCACTT GTAGACTGGA CCCTTCATCA AGAGCACCAT GG

SBMu200 Promoter Analysis: 5' Deletions

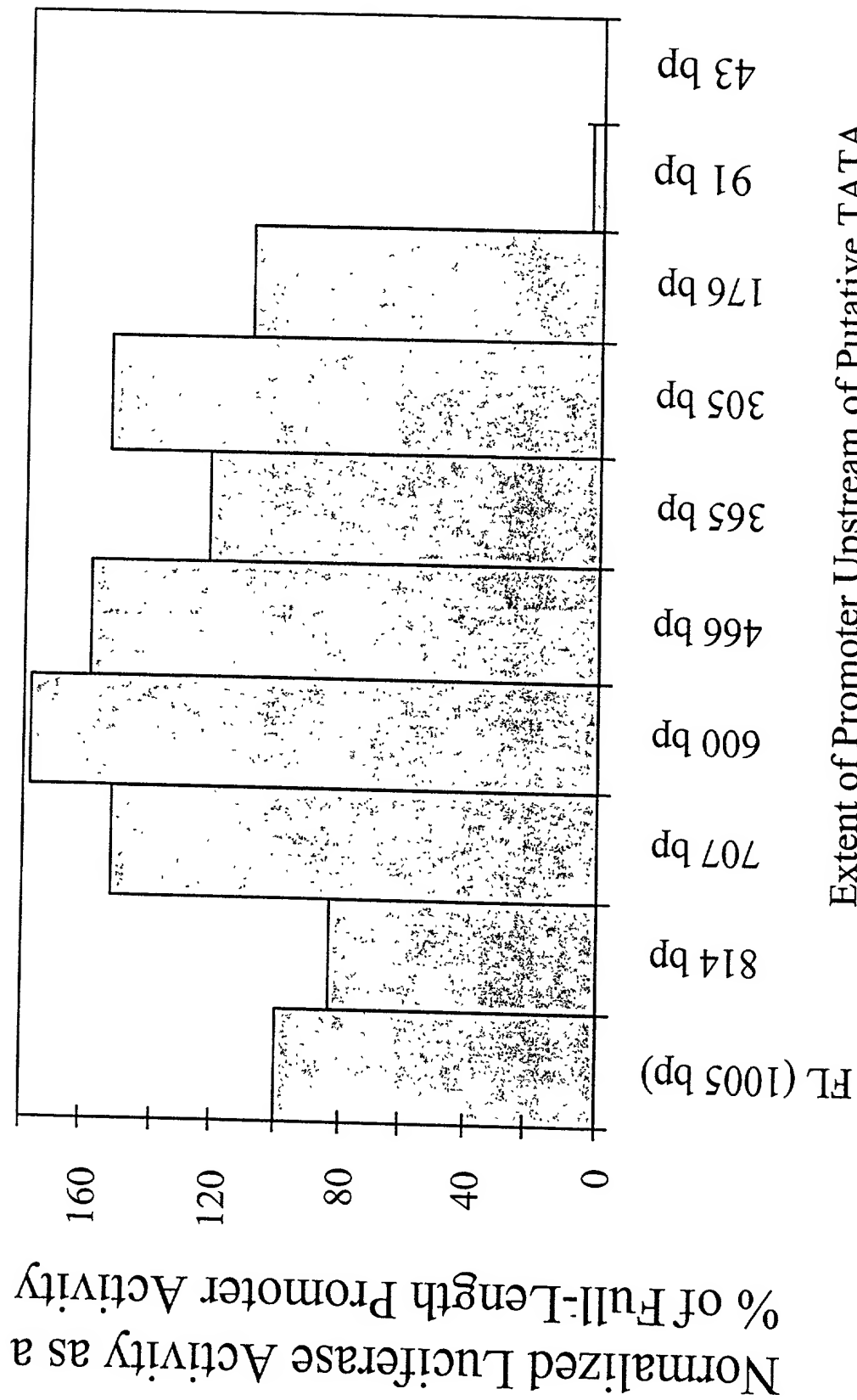


Figure 9

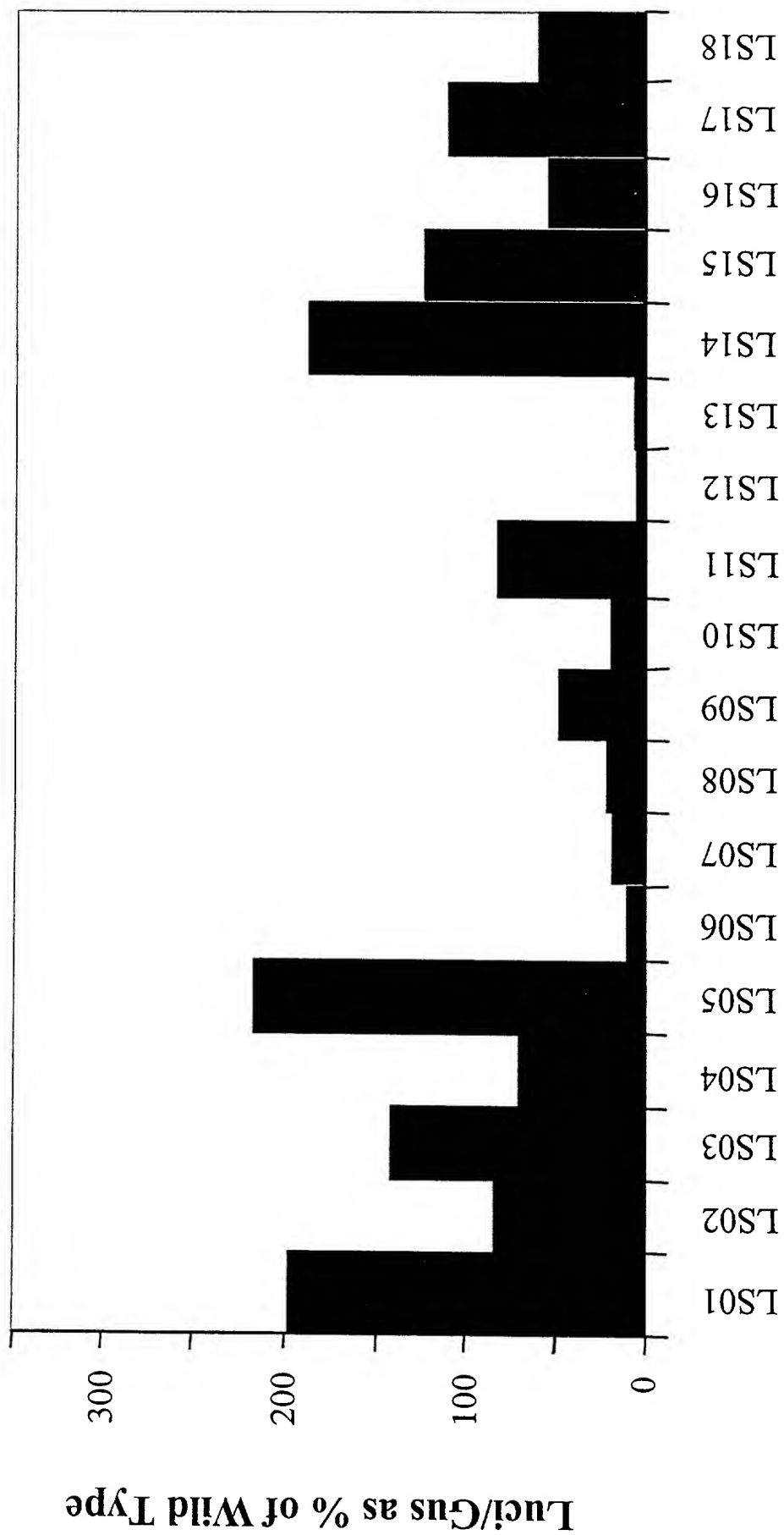
SBMu200 “Minimal” Promoter

| | | | | | | | | | | |
|------|-------------------|---------------|-------------|-------------------|-------------|-------------------|------|--|-------------|--|
| -180 | CCC | <u>ATCTCA</u> | TTTTCTTGGT | TGGCAGATCA | CAAAAAGGAA | CACAAAGGCT | | | | |
| | LS01 | | LS02 | | LS03 | | LS04 | | LS05 | |
| -130 | <u>AAGCCTCCTA</u> | CTTGTT | CGGG | AGTTAGGTCA | GGGACACCAT | <u>ATGAATGAAA</u> | | | | |
| | <i>LS06</i> | | <i>LS07</i> | | <i>LS08</i> | | LS09 | | <i>LS10</i> | |
| -80 | <u>GAAATCTTAA</u> | TTTGGGGTCA | CACCAAGATT | GTCCTCTG | TCG | <u>AGGTTGGGGG</u> | | | | |
| | LS11 | | <i>LS12</i> | | <i>LS13</i> | | LS14 | | LS15 | |
| -30 | <u>GTCCCTAAGG</u> | TTGGTAGTAG | CAATACCCAA | <u>TATATCACCT</u> | AACAAACCCA | | | | | |
| | LS16 | | LS17 | | LS18 | | | | | |
| 20 | ATCCATGCTA | CATACATACA | TAGCATCCAT | CAC TTGTAGA | CTGGACCCTT | | | | | |
| 70 | CATCAAGAGC | ACCATGG | | | | | | | | |

Coordinates are relative to the putative TATA box (underlined). P motifs are in *italic*.
 [] = Del -176/-92 [] = Del -89/-44 [] = Del -39/-8
 Linker scanning mutations that reduce activity to ~5% or less are in **bold**. Mutations with a significant but less pronounced effect are in **bold italic**.

Figure 10

Linker Scanning Analysis of SBMu200 Promoter



Mutant

Figure 11

[illegible]

Percent Similarity: 92.510 Percent Identity: 90.891
Sb200-Sorqhr.Pep x Sb20081.Pep February 13, 1997 11:29 ..

5 MPFTSYTYIADPVNVEHVLKTNFTNYPKGDVYRSYMDVLLGDGIFNADGE 54
87 MPFTSYTYIADPVNVEHVLKTNFTNYPKGIVYRSYMDVLLGDGIFNADGE 136
55 LWRKQRKTASFEFASKNLRDFSANVFREYSLKLSGILSQASKAGKVVDQM 104
137 LWRKQRKTASFEFASKNLRDFS AIVFREYSLKLSGILSQASKAGKVVDQM 186

Figure 12B

```

105 ELYMRMTLDSICXVGFGVXI GTLS PDL PEN SFXQAFDAANIIVTLRFIHP 154
    ||||| | | | | | | | | | | | | | | | | : | | | . |
187 ELYMRMTLDSICKVGFGVEIGT LSPDLPENSFAQAFDAANIIITLRFIDP 236

155 LWRIQKFF 162
    |||.:.|
237 LWRIKRFF 244

```

[illegible]